

When is a Function Securely Computable?

Himanshu Tyagi, Prakash Narayan and Piyush Gupta

Abstract

A subset of a set of terminals that observe correlated signals seek to compute a given function of the signals using public communication. It is required that the value of the function be kept secret from an eavesdropper with access to the communication. We show that the function is securely computable if and only if its entropy is less than the “aided secret key” capacity of an associated secrecy generation model, for which a single-letter characterization is provided.

Index Terms

Aided secret key, balanced coloring lemma, function computation, maximum common function, omniscience, secret key capacity, secure computability.

I. INTRODUCTION

In an online auction, $m - 1$ bidders acting independently of each other, randomly place one of k bids on a secure server. After a period of independent daily bidding, the server posts a cryptic message on a public website. Our results show that for $m > k + 1$, such a message exists from which each bidder can deduce securely the highest bids, but no message exists to allow any of them to identify securely the winners.

In general, suppose that the terminals in $\mathcal{M} = \{1, \dots, m\}$ observe correlated signals, and that a subset $\mathcal{A} = \{1, \dots, a\}$ of them are required to compute “securely” a given (single-letter) function g of all the signals. To this end, following their observations, all the terminals are allowed to communicate interactively over a public noiseless channel of unlimited capacity, with all such communication being observed by all the terminals. The terminals in \mathcal{A} seek to compute g in such a manner as to keep its value information theoretically secret from an eavesdropper with access to the public interterminal communication. See Figure 1. A typical application arises in a wireless network of colocated sensors which seek to compute a given function of their correlated measurements using public communication that does not give away the value of the function.

Our goal is to characterize necessary and sufficient conditions under which such secure computation is feasible. We formulate a new Shannon theoretic multiterminal source model that addresses the elemental question: *When can a function g be computed so that its value is independent of the public communication used in its computation?*

The work of H. Tyagi and P. Narayan was supported by the U.S. National Science Foundation under Grants CCF0635271 and CCF0830697. P. Gupta acknowledges support from NSF Grant CNS-519535.

H. Tyagi and P. Narayan are with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742, USA. E-mail: tyagi, prakash@umd.edu

P. Gupta is with Bell Labs, Alcatel-Lucent, Murray Hill, NJ 07974, USA. E-mail: pgupta@research.bell-labs.com

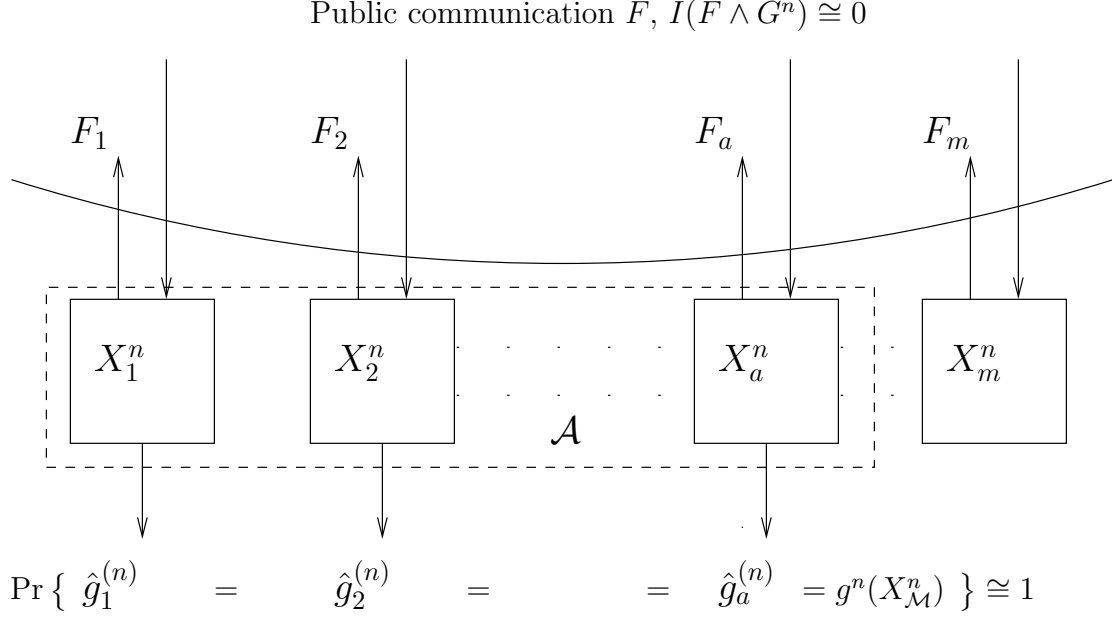


Fig. 1. Secure computation of g

We establish that the answer to this question is innately connected to a problem of secret key (SK) generation in which all the terminals in \mathcal{M} seek to generate “secret common randomness” at the largest rate possible, when the terminals in $\mathcal{A}^c = \mathcal{M}/\mathcal{A}$ are provided with side information for limited use, by means of public communication from which an eavesdropper can glean only a negligible amount of information about the SK. The public communication from a terminal can be any function of its own observed signal and of all previous communication. Side information is provided to the terminals in \mathcal{A}^c in the form of the value of g , and can be used only for recovering the key. Such a key, termed an aided secret key (ASK), constitutes a modification of the original notion of a SK in [14], [1], [6], [7]. The largest rate of such an ASK, which can be used for encrypted communication, is the ASK capacity C . Since a securely computable function g for \mathcal{A} will yield an ASK (for \mathcal{M}) of rate equal to its entropy H , it is clear that g necessarily must satisfy $H \leq C$. We show that surprisingly, $H < C$ is a sufficient condition for the existence of a protocol for the secure computation of g for \mathcal{A} . When all the terminals in \mathcal{M} seek to compute g securely, the corresponding ASK capacity reduces to the standard SK capacity for \mathcal{M} [6], [7]. We also show that a function that is securely computed by \mathcal{A} can be augmented by residual secret common randomness to yield a SK for \mathcal{A} of optimum rate.

We also present the capacity for a general ASK model involving *arbitrary* side information at the secrecy-seeking set of terminals for key recovery alone. Its capacity is characterized in terms of the classic concept of “maximum common function” [8]. Although this result is not needed in full dose for characterizing secure computability, it remains of independent interest.

We do not tackle the difficult problem of determining the minimum rate of public communication needed for the secure computation of g , which remains open even in the absence of a secrecy constraint

[11]. Nor do we fashion efficient protocols for this purpose. Instead, our mere objective in this work is to find conditions for the *existence* of such protocols.

The study of problems of function computation, with and without secrecy requirements, has a long and varied history to which we can make only a skimpy allusion here. Examples include: algorithms for exact function computation by multiple parties (cf. e.g., [20], [9], [10]); algorithms for asymptotically accurate (in observation length) function computation (cf. e.g., [18], [13]); exact function computation with secrecy (cf. e.g., [17]); and problems of oblivious transfer [16], [2].

Our results in Section III are organized in three parts: capacity of ASK model; characterization of the secure computability of g ; and a decomposition result for the total entropy of the model. Proofs are provided in Section IV and concluding remarks in Section V.

II. PRELIMINARIES

Let X_1, \dots, X_m , $m \geq 2$, be rvs with finite alphabets $\mathcal{X}_1, \dots, \mathcal{X}_m$, respectively. For any nonempty set $A \subseteq \mathcal{M} = \{1, \dots, m\}$, we denote $X_A = (X_i, i \in A)$. Similarly, for real numbers R_1, \dots, R_m and $A \subseteq \mathcal{M}$, we denote $R_A = (R_i, i \in A)$. Let A^c be the set $\mathcal{M} \setminus A$. We denote n i.i.d. repetitions of $X_{\mathcal{M}} = (X_1, \dots, X_m)$ with values in $\mathcal{X}_{\mathcal{M}} = \mathcal{X}_1 \times \dots \times \mathcal{X}_m$ by $X_{\mathcal{M}}^n = (X_1^n, \dots, X_m^n)$ with values in $\mathcal{X}_{\mathcal{M}}^n = \mathcal{X}_1^n \times \dots \times \mathcal{X}_m^n$. Following [6], given $\epsilon > 0$, for rvs U, V , we say that U is ϵ -recoverable from V if $\Pr(U \neq f(V)) \leq \epsilon$ for some function $f(V)$ of V . All logarithms and exponentials are with respect to the base 2.

We consider a multiterminal source model for secure computation with public communication; this basic model was introduced in [6] in the context of SK generation with public transaction. Terminals $1, \dots, m$ observe, respectively, the sequences X_1^n, \dots, X_m^n , of length n . Let $g : \mathcal{X}_{\mathcal{M}} \rightarrow \mathcal{Y}$ be a given mapping, where \mathcal{Y} is a finite alphabet. For $n \geq 1$, the mapping $g^n : \mathcal{X}_{\mathcal{M}}^n \rightarrow \mathcal{Y}^n$ is defined by

$$g^n(x_{\mathcal{M}}^n) = (g(x_{11}, \dots, x_{m1}), \dots, g(x_{1n}, \dots, x_{mn})),$$

$$x_{\mathcal{M}}^n = (x_1^n, \dots, x_m^n) \in \mathcal{X}_{\mathcal{M}}^n.$$

For convenience, we shall denote the rv $g^n(X_{\mathcal{M}}^n)$ by G^n , $n \geq 1$, and, in particular, $G^1 = g(X_{\mathcal{M}})$ simply by G . The terminals in a given set $\mathcal{A} \subseteq \mathcal{M}$ wish to “compute securely” the function $g^n(x_{\mathcal{M}}^n)$ for $x_{\mathcal{M}}^n$ in $\mathcal{X}_{\mathcal{M}}^n$. To this end, the terminals are allowed to communicate over a noiseless public channel, possibly interactively in several rounds. Randomization at the terminals is permitted; we assume that terminal i generates a rv U_i , $i \in \mathcal{M}$, such that U_1, \dots, U_m and $X_{\mathcal{M}}^n$ are mutually independent. While the cardinalities of range spaces of $U_i, i \in \mathcal{M}$, are unrestricted, we assume that $H(U_{\mathcal{M}}) < \infty$.

Definition 1. Assume without any loss of generality that the communication of the terminals in \mathcal{M} occurs in consecutive time slots in r rounds; such communication is described in terms of the mappings

$$f_{11}, \dots, f_{1m}, f_{21}, \dots, f_{2m}, \dots, f_{r1}, \dots, f_{rm},$$

with f_{ji} corresponding to a message in time slot j by terminal i , $1 \leq j \leq r$, $1 \leq i \leq m$; in general, f_{ji} is allowed to yield any function of (U_i, X_i^n) and of previous communication described in terms of $\{f_{kl} : k < j, l \in \mathcal{M} \text{ or } k = j, l < i\}$. The corresponding rvs representing the communication will be

depicted collectively as

$$\mathbf{F} = \{F_{11}, \dots, F_{1m}, F_{21}, \dots, F_{2m}, \dots, F_{r1}, \dots, F_{rm}\},$$

where $\mathbf{F} = \mathbf{F}^{(n)}(U_{\mathcal{M}}, X_{\mathcal{M}}^n)$. A special form of such communication will be termed *noninteractive communication* if $\mathbf{F} = (F_1, \dots, F_m)$, where $F_i = f_i(X_i^n)$, $i \in \mathcal{M}$.

Definition 2. For $\epsilon_n > 0, n \geq 1$, we say that g is ϵ_n -securely computable (ϵ_n -SC) by (the terminals in) a given set $\mathcal{A} \subseteq \mathcal{M}$ with $|\mathcal{A}| \geq 1$ from observations of length n , randomization $U_{\mathcal{M}}$ and public communication $\mathbf{F} = \mathbf{F}^{(n)}$, if

(i) g^n is ϵ_n -recoverable from (U_i, X_i^n, \mathbf{F}) for every $i \in \mathcal{A}$, i.e., there exists $\hat{g}_i^{(n)}$ satisfying

$$\Pr \left(\hat{g}_i^{(n)}(U_i, X_i^n, \mathbf{F}) \neq G^n \right) \leq \epsilon_n, \quad i \in \mathcal{A}, \quad (1)$$

and

(ii) g^n satisfies the “strong” secrecy condition¹

$$I(G^n \wedge \mathbf{F}) \leq \epsilon_n. \quad (2)$$

By definition, an ϵ_n -SC function g is recoverable (as g^n) at the terminals in \mathcal{A} and is effectively concealed from an eavesdropper with access to the public communication \mathbf{F} .

Definition 3. We say that g is *securely computable* by \mathcal{A} if g is ϵ_n -SC by \mathcal{A} from observations of length n , suitable randomization $U_{\mathcal{M}}$ and public communication \mathbf{F} , such that $\lim_n \epsilon_n = 0$.

III. WHEN IS g SECURELY COMPUTABLE?

We consider first the case when all the terminals in \mathcal{M} wish to compute securely the function g , i.e., $\mathcal{A} = \mathcal{M}$. Our result for this case will be seen to be linked inherently to the standard concept of SK capacity for a multiterminal source model [6], [7], and serves to motivate our approach to the general case when $\mathcal{A} \subseteq \mathcal{M}$.

Definition 4. [6], [7] For $\epsilon_n > 0, n \geq 1$, a function K of $(U_{\mathcal{M}}, X_{\mathcal{M}}^n)$ is an ϵ_n -secret key (ϵ_n -SK) for (the terminals in) a given set² $\mathcal{A}' \subseteq \mathcal{M}$ with $|\mathcal{A}'| \geq 2$, achievable from observations of length n , randomization $U_{\mathcal{M}}$ and public communication $\mathbf{F} = \mathbf{F}^{(n)}(U_{\mathcal{M}}, X_{\mathcal{M}}^n)$ as above if

- (i) K is ϵ_n -recoverable from (U_i, X_i^n, \mathbf{F}) for every $i \in \mathcal{A}'$;
- (ii) K satisfies the “strong” secrecy condition

$$\log |\mathcal{K}| - H(K | \mathbf{F}) = \log |\mathcal{K}| - H(K) + I(K \wedge \mathbf{F}) \leq \epsilon_n, \quad (3)$$

where $\mathcal{K} = \mathcal{K}^{(n)}$ denotes the set of possible values of K . The SK capacity $C(\mathcal{A}')$ for \mathcal{A}' is the largest rate $\lim_n (1/n) \log |\mathcal{K}^{(n)}|$ of ϵ_n -SKs for \mathcal{A}' as above, such that $\lim_n \epsilon_n = 0$.

¹The notion of strong secrecy for SK generation was introduced in [15], and developed further in [4], [5].

²For reasons of notation that will be apparent later, we distinguish between the secrecy seeking set $\mathcal{A}' \subseteq \mathcal{M}$ and the set $\mathcal{A} \subseteq \mathcal{M}$ pursuing secure computation.

Remarks. (i) The secrecy condition (3) is tantamount jointly to a nearly uniform distribution for K (i.e., $\log |\mathcal{K}| - H(K)$ is small) and to the near independence of K and \mathbf{F} (i.e., $I(K \wedge \mathbf{F})$ is small).

(ii) For the trivial case $|\mathcal{A}'| = 1$, clearly $C(\mathcal{A}') = H(X_{\mathcal{A}'})$.

A single-letter characterization of the SK capacity $C(\mathcal{A}')$ is provided in [6], [7].

Theorem 1. [6], [7] *The SK capacity $C(\mathcal{A}')$ equals*

$$C(\mathcal{A}') = H(X_{\mathcal{M}}) - R_{CO}(\mathcal{A}'), \quad (4)$$

where

$$R_{CO}(\mathcal{A}') = \min_{R_{\mathcal{M}} \in \mathcal{R}(\mathcal{A}')} \sum_{i=1}^m R_i \quad (5)$$

with

$$\mathcal{R}(\mathcal{A}') = \left\{ R_{\mathcal{M}} : R_B \geq H(X_B | X_{B^c}), \quad B \subsetneq \mathcal{M}, \mathcal{A}' \not\subseteq B \right\}. \quad (6)$$

Furthermore, the SK capacity can be achieved with noninteractive communication and without recourse to randomization at the terminals in \mathcal{M} .

Remark. The SK capacity $C(\mathcal{A}')$ is not increased if the secrecy condition (3) is replaced by either of the following weaker requirements³ [14], [6]:

$$\frac{1}{n} I(K \wedge \mathbf{F}) \leq \epsilon_n \quad \text{and} \quad \frac{1}{n} (\log |\mathcal{K}| - H(K)) \leq \epsilon_n, \quad (7)$$

or

$$\frac{1}{n} I(K \wedge \mathbf{F}) \leq \epsilon_n \quad \text{and} \quad \limsup_n \frac{1}{n} \log |\mathcal{K}| < \infty. \quad (8)$$

We recall from [6] that $R_{CO}(\mathcal{A}')$ has the operational significance of being the smallest rate of “communication for omniscience” for \mathcal{A}' , namely the smallest rate $\lim_n (1/n) \log \|\mathbf{F}^{(n)}\|$ of suitable communication for the terminals in \mathcal{M} whereby $X_{\mathcal{M}}^n$ is ϵ_n -recoverable from $(U_i, X_i^n, \mathbf{F}^n)$ at each terminal $i \in \mathcal{A}'$, with $\lim_n \epsilon_n = 0$; here $\|\mathbf{F}^{(n)}\|$ denotes the cardinality of the set of values of $\mathbf{F}^{(n)}$. Thus, $R_{CO}(\mathcal{A}')$ is the smallest rate of interterminal communication among the terminals in \mathcal{M} that enables every terminal in \mathcal{A}' to reconstruct with high probability all the sequences observed by all the other terminals in \mathcal{M} with the cooperation of the terminals in \mathcal{M}/\mathcal{A}' . The resulting omniscience for \mathcal{A}' corresponds to total “common randomness” of rate $H(X_{\mathcal{M}})$. The notion of omniscience, which plays a central role in SK generation for the multiterminal source model [6], will play a material role in the secure computation of g as well.

Noting that $g^n : \mathcal{X}_{\mathcal{M}}^n \rightarrow \mathcal{Y}^n$ implies

$$\frac{1}{n} \log |g^n(\mathcal{X}_{\mathcal{M}}^n)| \leq \log |\mathcal{X}_{\mathcal{M}}|, \quad (9)$$

³When randomization at the terminals in \mathcal{M} is not permitted, the converse proof in [6] uses only the first part of (7) or (8). When randomization is allowed, since the cardinality of the range space of $U_{\mathcal{M}}$ is unrestricted, the converse proof in [6] uses additionally the second part of (7) or (8).

a comparison of the conditions in (2, 9) and (8) that must be met by a securely computable g and a SK K , respectively, shows for a given g to be securely computable, it is necessary that

$$H(G) \leq C(\mathcal{M}). \quad (10)$$

Remarkably, it transpires that $H(G) < C(\mathcal{M})$ is a sufficient condition for g to be securely computable, and constitutes our first result.

Theorem 2. *A function g is securely computable by \mathcal{M} if*

$$H(G) < C(\mathcal{M}). \quad (11)$$

Conversely, if g is securely computable by \mathcal{M} , then $H(G) \leq C(\mathcal{M})$.

Theorem 2 is, in fact, a special case of our main result in Theorem 5 below.

Example 1. Let $m = 2$, and let X_1 and X_2 be $\{0, 1\}$ -valued rvs with

$$\begin{aligned} P_{X_1}(1) &= p = 1 - P_{X_1}(0), \quad 0 < p < 1, \\ P_{X_2|X_1}(1 | 1) &= P_{X_2|X_1}(0 | 0) = 1 - \delta, \quad 0 < \delta < \frac{1}{2}. \end{aligned}$$

Let $g(x_1, x_2) = x_1 + x_2 \pmod{2}$.

From [14], [1] (and also Theorem 1 above), $C(\{1, 2\}) = h(p * \delta) - h(\delta)$, where $p * \delta = (1 - p)\delta + p(1 - \delta)$. Since $H(G) = h(\delta)$, by Theorem 2 g is securely computable if

$$2h(\delta) < h(p * \delta). \quad (12)$$

We give a simple scheme for the secure computation of g when $p = \frac{1}{2}$, that relies on Wyner's well-known method for Slepian-Wolf data compression [19] and a derived SK generation scheme in [22], [21]. We can write

$$X_1^n = X_2^n + G^n \pmod{2} \quad (13)$$

with G^n being independent separately of X_2^n and X_1^n . We observe as in [19] that there exists a binary linear code, of rate $\cong 1 - h(\delta)$, with parity check matrix \mathbf{P} such that X_1^n , and so G^n , is ϵ_n -recoverable from (F_1, X_2^n) at terminal 2, where the Slepian-Wolf codeword $F_1 = \mathbf{P}X_1^n$ constitutes public communication from terminal 1, and where ϵ_n decays to 0 exponentially rapidly in n . Let \widehat{G}^n be the estimate of G^n thereby formed at terminal 2. Further, let $K = K(X_1^n)$ be the location of X_1^n in the coset of the standard array corresponding to \mathbf{P} . By the previous observation, K too is ϵ_n -recoverable from (F_1, X_2^n) at terminal 2. From [22], [21], K constitutes a “perfect” SK for terminals 1 and 2, of rate $\cong I(X_1 \wedge X_2) = 1 - h(\delta)$, and satisfying

$$I(K \wedge F_1) = 0. \quad (14)$$

Also, observe from (13) that $K = K(X_1^n) = K(X_2^n + G^n)$ and $F_1 = F_1(X_1^n) = F_1(X_2^n + G^n)$, and for each fixed value of G^n , the (common) arguments of K and F_1 have the same distribution as X_1^n .

Hence by (14),

$$I(K \wedge F_1, G^n) = I(K \wedge F_1 \mid G^n) = 0, \quad (15)$$

since $I(K \wedge G^n) \leq I(X_1^n \wedge G^n) = 0$.

Then terminal 2 communicates \widehat{G}^n in encrypted form as

$$F_2 = \widehat{G}^n + K \pmod{2}$$

(all represented in bits), with encryption feasible since

$$H(G) = h(\delta) < 1 - h(\delta) \cong \frac{1}{n} H(K),$$

by the sufficient condition (12). Terminal 1 then decrypts F_2 using K to recover \widehat{G}^n . The computation of g^n is secure since

$$I(G^n \wedge F_1, F_2) = I(G^n \wedge F_1) + I(G^n \wedge F_2 \mid F_1)$$

is small; specifically, the first term equals 0 since $I(G^n \wedge F_1) \leq I(G^n \wedge X_1^n) = 0$, while the second term is bounded using (15) according to

$$\begin{aligned} I(G^n \wedge F_2 \mid F_1) &= H(\widehat{G}^n + K \mid F_1) - H(\widehat{G}^n + K \mid F_1, G^n) \\ &\leq H(K) - H(G^n + K \mid F_1, G^n) + \delta_n \\ &= I(K \wedge F_1, G^n) + \delta_n = \delta_n, \end{aligned}$$

where the inequality follows by Fano's inequality and the exponential decay of ϵ_n to 0. \square

Next, we turn to the general model for the secure computability of g by a given set $\mathcal{A} \subseteq \mathcal{M}$. Again in the manner of (10), it is clear that a necessary condition is

$$H(G) \leq C(\mathcal{A}).$$

In contrast, when $\mathcal{A} \subsetneq \mathcal{M}$, $H(G) < C(\mathcal{A})$ is *not* sufficient for g to be securely computable by \mathcal{A} as seen by the following simple example.

Example 2. Let $m = 3$, $A = \{1, 2\}$ and consider rvs X_1, X_2, X_3 with $X_1 = X_2$, where X_1 is independent of X_3 and $H(X_3) < H(X_1)$. Let g be defined by $g(x_1, x_2, x_3) = x_3$, $x_i \in \mathcal{X}_i$, $1 \leq i \leq 3$. Clearly, $C(\{1, 2\}) = H(X_1)$. Therefore, $H(G) = H(X_3) < C(\{1, 2\})$. However, for g to be computed by the terminals 1 and 2, its value must be conveyed to them necessarily by public communication from terminal 3. Thus, g is not securely computable. \square

Interestingly, the secure computability of g can be examined in terms of a new SK generation problem that is formulated next.

A. Secret Key Aided by Side Information

We consider an extension of the SK generation problem in Definition 4, which involves additional side information $Z_{\mathcal{A}'}^n$ that is correlated with $X_{\mathcal{M}}^n$ and is provided to the terminals in \mathcal{A}' for use in *only*

the recovery stage of SK generation; however, the public communication \mathbf{F} remains as in Definition 1. Formally, the extension is described in terms of generic rvs $(X_1, \dots, X_m, \{Z_i, i \in \mathcal{A}'\})$, where the rvs Z_i too take values in finite sets \mathcal{Z}_i , $i \in \mathcal{A}'$. We note that the full force of this extension will not be needed to characterize the secure computability of g ; an appropriate particularization will suffice. Nevertheless, this concept is of independent interest.

Definition 5. A function K of $(U_{\mathcal{M}}, X_{\mathcal{M}}^n, Z_{\mathcal{A}'}^n)$ is an ϵ_n -secret key aided by side information $Z_{\mathcal{A}'}^n$ (ϵ_n -ASK) for the terminals $\mathcal{A}' \subseteq \mathcal{M}$, $|\mathcal{A}'| \geq 2$, achievable from observations of length n , randomization $U_{\mathcal{M}}$ and public communication $\mathbf{F} = \mathbf{F}(U_{\mathcal{M}}, X_{\mathcal{M}}^n)$ if it satisfies the conditions in Definition 4 with $(U_i, X_i^n, Z_i^n, \mathbf{F})$ in the role of (U_i, X_i^n, \mathbf{F}) in condition (i). The corresponding ASK capacity $C(\mathcal{A}', Z_{\mathcal{A}'})$ is defined analogously as in Definition 4.

In contrast with the omniscience rate of $H(X_{\mathcal{M}})$ that appears in the passage following Theorem 1, now an underlying analogous notion of omniscience will involve total common randomness of rate exceeding $H(X_{\mathcal{M}})$. Specifically, the enhanced common randomness rate will equal the entropy of the “maximum common function” (mcf) of the rvs $(X_{\mathcal{M}}, Z_i)_{i \in \mathcal{A}}$, introduced for a pair of rvs in [8] (see also [3, Problem 3.4.27]).

Definition 6. [8] For two rvs Q, R with values in finite sets \mathcal{Q}, \mathcal{R} , the equivalence relation $q \sim q'$ in \mathcal{Q} holds if there exist $N \geq 1$ and sequences (q_0, q_1, \dots, q_N) in \mathcal{Q} with $q_0 = q$, $q_N = q'$ and (r_1, \dots, r_N) in \mathcal{R} satisfying $\Pr(Q = q_{l-1}, R = r_l) > 0$ and $\Pr(Q = q_l, R = r_l) > 0$, $l = 1, \dots, N$. Denote the corresponding equivalence classes in \mathcal{Q} by $\mathcal{Q}_1, \dots, \mathcal{Q}_k$. Similarly, let $\mathcal{R}_1, \dots, \mathcal{R}_{k'}$ denote the equivalence classes in \mathcal{R} . As argued in [8], $k = k'$ and for $1 \leq i, j \leq k$,

$$\Pr(Q \in \mathcal{Q}_i \mid R \in \mathcal{R}_j) = \Pr(R \in \mathcal{R}_j \mid Q \in \mathcal{Q}_i) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

The mcf of the rvs Q, R is a rv $\text{mcf}(Q, R)$ with values in $\{1, \dots, k\}$ and pmf

$$\Pr(\text{mcf}(Q, R) = i) = \Pr(Q \in \mathcal{Q}_i) = \Pr(Q \in \mathcal{Q}_i, R \in \mathcal{R}_i), \quad i = 1, \dots, k.$$

For rvs Q_1, \dots, Q_m taking values in finite alphabets, we define the $\text{mcf}(Q_1, \dots, Q_m)$ recursively by

$$\text{mcf}(Q_1, \dots, Q_m) = \text{mcf}(\text{mcf}(Q_1, \dots, Q_{m-1}), Q_m) \quad (16)$$

with $\text{mcf}(Q_1, Q_2)$ as above.

Definition 7. With Q^n denoting n i.i.d. repetitions of the rv Q , we define

$$\text{mcf}^n(Q_1, \dots, Q_m) = \{\text{mcf}(Q_{1t}, \dots, Q_{mt})\}_{t=1}^n. \quad (17)$$

Note that $\text{mcf}^n(Q_1, \dots, Q_m)$ is a function of *each* individual $Q_i^n, i = 1, \dots, m$.

Remark. As justification for the definition (16), consider a rv ξ that satisfies

$$H(\xi \mid Q_i) = 0, \quad i = 1, \dots, m \quad (18)$$

and suppose for any other rv ξ' satisfying (18) that $H(\xi) \geq H(\xi')$. Then Lemma 3 below shows that ξ must satisfy $H(\xi) = H(\text{mcf}(Q_1, \dots, Q_m))$.

The following result for the mcf of $m \geq 2$ rvs is a simple extension of the classic result for $m = 2$ [8, Theorem 1].

Lemma 3. *Given $0 < \epsilon < 1$, if $\xi^{(n)}$ is ϵ -recoverable from Q_i^n for each $i = 1, \dots, m$, then*

$$\limsup_n \frac{1}{n} H(\xi^{(n)}) \leq H(\text{mcf}(Q_1, \dots, Q_m)). \quad (19)$$

Proof: The proof involves a recursive application of [8, Lemma, Section 4] to $\text{mcf}(Q_1, \dots, Q_m)$ in (16), and is provided in Appendix A.

We are now in a position to characterize ASK capacity. In a manner analogous to Theorem 1, this is done in terms of $H(\text{mcf}(X_{\mathcal{M}}, Z_i)_{i \in \mathcal{A}'})$ and the smallest rate of communication $R_{CO}(\mathcal{A}', Z_{\mathcal{A}'})$ for each terminal in \mathcal{A}' to attain omniscience that corresponds to n i.i.d. repetitions of $\text{mcf}(X_{\mathcal{M}}, Z_i)_{i \in \mathcal{A}'}$.

Theorem 4. *The ASK capacity $C(\mathcal{A}'; Z_{\mathcal{A}'})$ equals*

$$C(\mathcal{A}'; Z_{\mathcal{A}'}) = H(\text{mcf}((X_{\mathcal{M}}, Z_i)_{i \in \mathcal{A}'})) - R_{CO}(\mathcal{A}'; Z_{\mathcal{A}'})$$

where

$$R_{CO}(\mathcal{A}'; Z_{\mathcal{A}'}) = \min_{R_{\mathcal{M}} \in \mathcal{R}(\mathcal{A}'; Z_{\mathcal{A}'})} \sum_{i \in \mathcal{M}} R_i$$

with

$$\mathcal{R}(\mathcal{A}'; Z_{\mathcal{A}'}) = \left\{ R_{\mathcal{M}} : R_B \geq \max_{j \in B^c \cap \mathcal{A}'} H(X_B | X_{B^c}, Z_j), \quad B \subsetneq \mathcal{M}, \mathcal{A}' \not\subseteq B \right\}. \quad (20)$$

The proof of Theorem 4 is along the same lines as that of Theorem 1 [6] and is provided in Appendix B.

The remark following Theorem 1 also applies to the ASK capacity $C(\mathcal{A}'; Z_{\mathcal{A}'})$, as will be seen from the proof of Theorem 4.

B. Characterization of Secure Computability

If g is securely computable by the terminals in \mathcal{A} , then G^n constitutes an ASK for \mathcal{M} under the constraint (8), of rate $H(G)$, with side information in the form of G^n provided only to the terminals in \mathcal{A}^c in the recovery stage of SK generation. Thus, a necessary condition for g to be securely computable by \mathcal{A} , in the manner of (10), is

$$H(G) \leq C(\mathcal{M}; Z_{\mathcal{M}}), \quad (21)$$

where $Z_{\mathcal{M}} = Z_{\mathcal{M}}(\mathcal{A}) = \{Z_i\}_{i \in \mathcal{M}}$ with

$$Z_i = \begin{cases} 0, & i \in \mathcal{A} \\ G, & i \in \mathcal{A}^c. \end{cases} \quad (22)$$

By particularizing Theorem 4 to the choice of $Z_{\mathcal{M}}$ as above, the right side of (21) reduces to

$$C(\mathcal{M}; Z_{\mathcal{M}}) = H(X_{\mathcal{M}}) - R_{CO}(\mathcal{M}; Z_{\mathcal{M}}) \quad (23)$$

where

$$R_{CO}(\mathcal{M}; Z_{\mathcal{M}}) = \min_{R_{\mathcal{M}} \in \mathcal{R}(\mathcal{M}; Z_{\mathcal{M}})} \sum_{i \in \mathcal{M}} R_i$$

with

$$\mathcal{R}(\mathcal{M}; Z_{\mathcal{M}}) = \left\{ R_{\mathcal{M}} : R_B \geq \begin{cases} H(X_B | X_{B^c}), & B \subsetneq \mathcal{M}, \mathcal{A} \not\subseteq B \\ H(X_B | X_{B^c}, G), & B \subsetneq \mathcal{M}, \mathcal{A} \subseteq B \end{cases} \right\}.$$

Our main result says that the necessary condition (21) is tight.

Theorem 5. *A function g is securely computable by $\mathcal{A} \subseteq \mathcal{M}$ if*

$$H(G) < C(\mathcal{M}; Z_{\mathcal{M}}). \quad (24)$$

Furthermore, under the condition above, g is securely computable with noninteractive communication and without recourse to randomization at the terminals in \mathcal{M} .

Conversely, if g is securely computable by $\mathcal{A} \subseteq \mathcal{M}$, then $H(G) \leq C(\mathcal{M}; Z_{\mathcal{M}})$.

Remarks. (i) It is easy to see that $C(\mathcal{M}) \leq C(\mathcal{M}; Z_{\mathcal{M}}) = C(\mathcal{M}; Z_{\mathcal{M}}(\mathcal{A})) \leq C(\mathcal{A})$. In particular, the second inequality holds since in the context of $C(\mathcal{M}; Z_{\mathcal{M}})$ the side information for recovery $Z_{\mathcal{M}}$ in (22) is not provided to the terminals in \mathcal{A} and by noting that a SK for \mathcal{M} is also a SK for \mathcal{A} .

(ii) Observe in Example 2 that $C(\mathcal{M}; Z_{\mathcal{M}}) = C(\mathcal{M}) = 0$ and so, by Theorem 5, g is not securely computable as noted earlier.

Example 3. For the auction example in Section I, $\mathcal{A} = \{1, \dots, m-1\}$ and X_1, \dots, X_{m-1} are i.i.d. rvs distributed uniformly on $\{1, \dots, k\}$, while $X_m = (X_1, \dots, X_{m-1})$. Let $g_1(x_1, \dots, x_m) = \max_{1 \leq i \leq m-1} x_i$ and $g_2(x_1, \dots, x_m) = \arg \max_{1 \leq i \leq m-1} x_i$. Then, straightforward computation yields for $k < m-1$ that

$$H(G_1) < \log k < H(G_2) = \log(m-1),$$

and for both g_1, g_2 that

$$C(\mathcal{M}; Z_{\mathcal{M}}) = C(\mathcal{M}),$$

where, by Theorem 1,

$$C(\mathcal{M}) = H(X_{\mathcal{M}}) - R_{CO}(\mathcal{M}) = (m-1) \log k - (m-2) \log k = \log k.$$

By Theorem 5, g_1 is securely computable whereas g_2 is not. In fact, g_2 is not securely computable by *any* terminal $i \in \{1, \dots, m-1\}$. This, too, is implied by Theorem 5 upon noting that for each $i \in \{1, \dots, m-1\}$ and a restricted choice $\mathcal{A} = \{i\}$,

$$C(\mathcal{M}; Z_{\mathcal{M}}(\mathcal{A})) = H(X_i) = \log k < \log(m-1) = H(G_2),$$

where the first equality is a consequence of remark (i) following Theorem 5 and remark (ii) after Definition 4. \square

C. A Decomposition Result

The sufficiency condition (24) prompts the following two natural questions: Does the difference $C(\mathcal{M}; Z_{\mathcal{M}}) - H(G)$ possess an operational significance? If g is securely computable by the terminals in \mathcal{A} , clearly G^n forms a SK for \mathcal{A} . Can G^n be augmented suitably to form a SK for \mathcal{A} of maximum achievable rate?

The answers to both these questions are in the affirmative. In particular, our approach to the second question involves a characterization of the minimum rate of communication for omniscience for \mathcal{A} , under the additional requirement that this communication be independent of G^n . Specifically, we show below that for a securely computable function g , this minimum rate remains $R_{CO}(\mathcal{A})$ (see (6)).

Addressing the first question, we introduce a rv $K_g = K_g^{(n)}$ such that $K = (K_g, G^n)$ constitutes an ϵ_n -ASK for \mathcal{M} with side information $Z_{\mathcal{M}}$ as in (22) and satisfying the additional requirement

$$I(K_g \wedge G^n) \leq \epsilon_n. \quad (25)$$

Let the largest rate $\lim_n (1/n) \log |\mathcal{K}_g^{(n)}|$ of such an ASK be $C^g(\mathcal{M}; Z_{\mathcal{M}})$. Observe that since K is required to be nearly independent of \mathbf{F} , where \mathbf{F} is the public communication involved in its formation, it follows by (25) that K_g is nearly independent of (G^n, \mathbf{F}) .

Turning to the second question, in the same vein let K'_g be a rv such that $K' = (K'_g, G^n)$ constitutes an ϵ_n -SK for $\mathcal{A} \subseteq \mathcal{M}$ and satisfying (25). Let $C^g(\mathcal{A})$ denote the largest rate of K'_g . As noted above, K'_g will be nearly independent of (G^n, \mathbf{F}') , where \mathbf{F}' is the public communication involved in the formation of K' .

Proposition 6. *For $\mathcal{A} \subseteq \mathcal{M}$, it holds that*

$$\begin{aligned} (i) \quad & C^g(\mathcal{M}; Z_{\mathcal{M}}(\mathcal{A})) = C(\mathcal{M}; Z_{\mathcal{M}}(\mathcal{A})) - H(G), \\ (ii) \quad & C^g(\mathcal{A}) = C(\mathcal{A}) - H(G). \end{aligned}$$

Remarks. (i) For the case $\mathcal{A} = \mathcal{M}$, both (i) and (ii) above reduce to $C^g(\mathcal{M}) = C(\mathcal{M}) - H(G)$.
(ii) Theorem 1 and Proposition 6(ii) lead to the observation

$$H(X_{\mathcal{M}}) = R_{CO}(\mathcal{A}) + H(G) + C^g(\mathcal{A}),$$

which admits the following heuristic interpretation. The “total randomness” $X_{\mathcal{M}}^n$ that corresponds to omniscience decomposes into three “nearly mutually independent” components: a minimum-sized communication for omniscience for \mathcal{A} and the independent parts of an optimum-rate SK for \mathcal{A} composed of G^n and K'_g .

IV. PROOFS OF THEOREM 5 AND PROPOSITION 6

A. Proof of Theorem 5

The necessity of (21) follows by the comments preceding Theorem 5.

The sufficiency of (24) will be established by showing the existence of *noninteractive* public communication comprising source codes that enable omniscience corresponding to $X_{\mathcal{M}}^n$ at the terminals

in \mathcal{A} , and thereby the computation of g . Furthermore, the corresponding codewords are selected so as to be simultaneously independent of G^n , thus assuring security.

First, from (24) and (23), there exists $\delta > 0$ such that $R_{CO}(\mathcal{M}; Z_{\mathcal{M}}) + \delta < H(X_{\mathcal{M}}|G)$, using $G = g(X_{\mathcal{M}})$. For each i and $R_i \geq 0$, consider a (map-valued) rv J_i that is uniformly distributed on the family \mathcal{J}_i of all mappings $\mathcal{X}_i^n \rightarrow \{1, \dots, \lceil \exp(nR_i) \rceil\}$, $i \in \mathcal{M}$. The rvs $J_1, \dots, J_m, X_{\mathcal{M}}^n$ are taken to be mutually independent.

Fix ϵ, ϵ' , with $\epsilon' > m\epsilon$ and $\epsilon + \epsilon' < 1$. It follows from the proof of the general source network coding theorem [3, Lemma 3.1.13 and Theorem 3.1.14] that for all sufficiently large n ,

$$\Pr \left(\left\{ j_{\mathcal{M}} \in \mathcal{J}_{\mathcal{M}} : X_{\mathcal{M}}^n \text{ is } \epsilon_n\text{-recoverable from } \left(X_i^n, j_{\mathcal{M} \setminus \{i\}} \left(X_{\mathcal{M} \setminus \{i\}}^n \right), Z_i^n \right), i \in \mathcal{M} \right\} \right) \geq 1 - \epsilon, \quad (26)$$

provided $R_{\mathcal{M}} = (R_1, \dots, R_m) \in \mathcal{R}(\mathcal{M}; Z_{\mathcal{M}})$, where ϵ_n vanishes exponentially rapidly in n . This assertion follows exactly as in the proof of [6, Proposition 1, with $A = \mathcal{M}$] but with \tilde{X}_i there equal to (X_i, Z_i) rather than X_i , $i \in \mathcal{M}$. In particular, we shall choose $R_{\mathcal{M}} \in \mathcal{R}(\mathcal{M}; Z_{\mathcal{M}})$ such that

$$\sum_{i=1}^m R_i \leq R_{CO}(\mathcal{M}; Z_{\mathcal{M}}) + \frac{\delta}{2}. \quad (27)$$

Below we shall establish that

$$\Pr \{ j_{\mathcal{M}} \in \mathcal{J}_{\mathcal{M}} : I(j_{\mathcal{M}}(X_{\mathcal{M}}^n) \wedge G^n) \geq \epsilon_n \} \leq \epsilon', \quad (28)$$

for all n sufficiently large, to which end it suffices to show that

$$\Pr \left(\left\{ j_{\mathcal{M}} \in \mathcal{J}_{\mathcal{M}} : I \left(j_i(X_i^n) \wedge G^n, j_{\mathcal{M} \setminus \{i\}} \left(X_{\mathcal{M} \setminus \{i\}}^n \right) \right) \geq \frac{\epsilon_n}{m} \right\} \leq \frac{\epsilon'}{m}, \quad i \in \mathcal{M}, \quad (29)$$

since

$$\begin{aligned} I(j_{\mathcal{M}}(X_{\mathcal{M}}^n) \wedge G^n) &= \sum_{i=1}^m I(j_i(X_i^n) \wedge G^n \mid j_1(X_1^n), \dots, j_{i-1}(X_{i-1}^n)) \\ &\leq \sum_{i=1}^m I(j_i(X_i^n) \wedge G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n)). \end{aligned}$$

Then it would follow from (26), (28) and definition of $Z_{\mathcal{M}}$ in (21) that

$$\Pr \left(\left\{ j_{\mathcal{M}} \in \mathcal{J}_{\mathcal{M}} : G^n \text{ is } \epsilon_n\text{-recoverable from } \left(X_i^n, j_{\mathcal{M} \setminus \{i\}} \left(X_{\mathcal{M} \setminus \{i\}}^n \right) \right), i \in \mathcal{A}, \right. \right. \\ \left. \left. \text{and } I(j_{\mathcal{M}}(X_{\mathcal{M}}^n) \wedge G^n) < \epsilon_n \right\} \right) \geq 1 - \epsilon - \epsilon'.$$

This shows the existence of a particular realization $j_{\mathcal{M}}$ of $J_{\mathcal{M}}$ such that G^n is ϵ_n -SC from $(X_i^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n))$ for each $i \in \mathcal{A}$.

It now remains to prove (29). Fix $i \in \mathcal{M}$ and note that for each $j_i \in \mathcal{J}_i$, with $\|j_i\|$ denoting the

cardinality of the (image) set $j_i(\mathcal{X}_i^n)$,

$$\begin{aligned} & I(j_i(X_i^n) \wedge G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n)) \\ & \leq I(j_i(X_i^n) \wedge G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n)) + \log \|j_i\| - H(j_i(X_i^n)) \\ & = D(j_i(X_i^n), (G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n)) \| U_{j_i(\mathcal{X}_i^n)} \times (G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n)), \end{aligned} \quad (30)$$

where the right side above denotes the (Kullback-Leibler) divergence between the joint pmf of $j_i(X_i^n), (G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n))$ and the product of the uniform pmf on $j_i(\mathcal{X}_i^n)$ and the pmf of $(G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n))$. Using [6, Lemma 1], the right side of (30) is bounded above further by

$$s_{var} \log \frac{\|j_i\|}{s_{var}}, \quad (31)$$

where $s_{var} = s_{var}(j_i(X_i^n); G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n))$ is the variational distance between the pmfs in the divergence above. Therefore, to prove (29), it suffices to show that

$$\Pr \left(\left\{ j_{\mathcal{M}} \in \mathcal{J}_{\mathcal{M}} : s_{var} \left(j_i(X_i^n); G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n) \right) \geq \frac{\epsilon_n}{m} \right\} \right) \leq \frac{\epsilon'}{m}, \quad i \in \mathcal{M}, \quad (32)$$

on account of the fact that $\log \|j_i(X_i^n)\| = O(n)$, and the exponential decay to 0 of ϵ_n . Defining

$$\tilde{\mathcal{J}}_i = \left\{ j_{\mathcal{M} \setminus \{i\}} \in \mathcal{J}_{\mathcal{M} \setminus \{i\}} : X_{\mathcal{M}}^n \text{ is } \epsilon_n\text{-recoverable from } (X_i^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n), Z_i^n) \right\},$$

we have by (26) that $\Pr(j_{\mathcal{M} \setminus \{i\}} \in \tilde{\mathcal{J}}_i) \geq 1 - \epsilon$. Thus, in (32),

$$\begin{aligned} & \Pr \left(\left\{ j_{\mathcal{M}} \in \mathcal{J}_{\mathcal{M}} : s_{var} \left(j_i(X_i^n); G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n) \right) \geq \frac{\epsilon_n}{m} \right\} \right) \\ & \leq \epsilon + \sum_{j_{\mathcal{M} \setminus \{i\}} \in \tilde{\mathcal{J}}_i} \Pr(j_{\mathcal{M} \setminus \{i\}} = j_{\mathcal{M} \setminus \{i\}}) \times \\ & \quad \Pr \left(\left\{ j_i \in \mathcal{J}_i : s_{var} \left(j_i(X_i^n); G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n) \right) \geq \frac{\epsilon_n}{m} \right\} \right), \end{aligned}$$

since J_i is independent of $J_{\mathcal{M} \setminus \{i\}}$. Thus, (32), and hence (29), will follow upon showing that

$$\Pr \left(\left\{ j_i \in \mathcal{J}_i : s_{var} \left(j_i(X_i^n); G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n) \right) \geq \frac{\epsilon_n}{m} \right\} \right) \leq \frac{\epsilon'}{m} - \epsilon, \quad j_{\mathcal{M} \setminus \{i\}} \in \tilde{\mathcal{J}}_i, \quad (33)$$

for all n sufficiently large. Fix $j_{\mathcal{M} \setminus \{i\}} \in \tilde{\mathcal{J}}_i$. We take recourse to Lemma C2 in Appendix C, and set $U = X_{\mathcal{M}}^n, U' = X_i^n, V = G^n, h = j_{\mathcal{M} \setminus \{i\}}$, and

$$\mathcal{U}_0 = \left\{ x_{\mathcal{M}}^n \in \mathcal{X}_{\mathcal{M}}^n : x_{\mathcal{M}}^n = \psi_i \left(x_i^n, j_{\mathcal{M} \setminus \{i\}}(x_{\mathcal{M} \setminus \{i\}}^n), g^n(x_{\mathcal{M}}^n) \mathbf{1}(i \in \mathcal{A}^c) \right) \right\}$$

for some mapping ψ_i . By the definition of $\tilde{\mathcal{J}}_i$,

$$\Pr(U \in \mathcal{U}_0) \geq 1 - \epsilon_n,$$

so that condition (C2)(i) preceding Lemma C2 is met. Condition (C2)(ii), too, is met since conditioned on the events in (C2)(ii), only those $x_{\mathcal{M}}^n \in \mathcal{U}_0$ can occur that are determined uniquely by their i^{th}

components x_i^n .

Upon choosing

$$d = \exp \left[n \left(H(X_{\mathcal{M}}|G) - \frac{\delta}{6} \right) \right],$$

in (C3), the hypotheses of Lemma C2 are satisfied with $\lambda = \sqrt{\epsilon_n}$, for an appropriate exponentially vanishing ϵ_n . Then, by Lemma C2, with

$$r = \lceil \exp[nR_i] \rceil, \quad r' = \left\lceil \exp \left[n \left(\sum_{l \in \mathcal{M} \setminus \{i\}} R_l + \frac{\delta}{6} \right) \right] \right\rceil,$$

and with J_i in the role of ϕ , we get from (C4) and (27) that

$$\Pr \left(\left\{ j_i \in \mathcal{J}_i : s_{var} \left(j_i(X_i^n); G^n, j_{\mathcal{M} \setminus \{i\}}(X_{\mathcal{M} \setminus \{i\}}^n) \right) \geq 14\sqrt{\epsilon_n} \right\} \right)$$

decays to 0 doubly exponentially in n , which proves (33). This completes the proof of Theorem 5. \square

B. Proof of Proposition 6

(i) Since the rv $(K_g^{(n)}, G^n)$, with nearly independent components, constitutes an ASK for \mathcal{M} with side information $Z_{\mathcal{M}}$ as in (22), it is clear that

$$H(G) + C^g(\mathcal{M}; Z_{\mathcal{M}}) \leq C(\mathcal{M}; Z_{\mathcal{M}}). \quad (34)$$

In order to prove the reverse of (34), we show that $C(\mathcal{M}; Z_{\mathcal{M}}) - H(G)$ is an achievable ASK rate for K_g that additionally satisfies (25). First, note that in the proof of Theorem 5, the assertions (26) and (29) mean that for all sufficiently large n , there exists a public communication $F_{\mathcal{M}}$, say, such that $I(F_{\mathcal{M}} \wedge G^n) < \epsilon_n$ and $X_{\mathcal{M}}^n$ is ϵ_n -recoverable from $(X_i^n, F_{\mathcal{M}}, Z_i^n)$ for every $i \in \mathcal{M}$, with $\lim_n \epsilon_n = 0$. Fix $0 < \tau < \delta$, where δ is as in the proof of Theorem 5. Apply Lemma C2, choosing

$$U = U' = X_{\mathcal{M}}^n, \quad \mathcal{U}_0 = \mathcal{X}_{\mathcal{M}}^n, \quad V = G^n, \quad h = F_{\mathcal{M}}, \quad d = \exp \left[n \left(H(X_{\mathcal{M}}|G) - \frac{\tau}{6} \right) \right], \quad (35)$$

whereby the hypothesis (C3) of Lemma C2 is satisfied for all n sufficiently large. Fixing

$$r' = \left\lceil \exp \left[n \left(R_{CO}(\mathcal{M}; Z_{\mathcal{M}}) + \frac{\tau}{2} \right) \right] \right\rceil,$$

by Lemma C2 a randomly chosen ϕ of rate

$$\frac{1}{n} \log r = H(X_{\mathcal{M}}|G) - R_{CO}(\mathcal{M}; Z_{\mathcal{M}}) - \tau = C(\mathcal{M}; Z_{\mathcal{M}}) - H(G) - \tau$$

will yield an ASK $K_g = K_g^{(n)} = \phi(X_{\mathcal{M}}^n)$ which is nearly independent of $(F_{\mathcal{M}}, G^n)$ (and, in particular, satisfies (25)) with positive probability, for all n sufficiently large.

(ii) The proof can be completed as that of part (i) upon showing that for a securely computable g , for all $\tau > 0$ and n sufficiently large, there exists a public communication $F'_{\mathcal{M}}$ that meets the following requirements: its rate does not exceed $R_{CO}(\mathcal{A}) + \tau$; $I(F'_{\mathcal{M}} \wedge G^n) < \epsilon_n$; and $X_{\mathcal{M}}^n$ is ϵ_n -recoverable from $(X_i^n, F'_{\mathcal{M}})$ for every $i \in \mathcal{A}$. To that end, for $R_{\mathcal{M}} = (R_1, \dots, R_m) \in \mathcal{R}(\mathcal{M}; Z_{\mathcal{M}})$ as in the proof

of Theorem 5, consider $R'_\mathcal{M} = (R'_1, \dots, R'_m) \in \mathcal{R}(\mathcal{A})$ that satisfies $R'_i \leq R_i$ for all $i \in \mathcal{M}$ and

$$\sum_{i=1}^m R'_i \leq R_{CO}(\mathcal{A}) + \tau,$$

noting that $\mathcal{R}(\mathcal{M}; Z_\mathcal{M}) \subseteq \mathcal{R}(\mathcal{A})$. Further, for $J_\mathcal{M}$ and $\mathcal{J}_\mathcal{M}$ as in that proof, define a (map-valued) rv J'_i that is uniformly distributed on the family \mathcal{J}'_i of all mappings from $\{1, \dots, \lceil \exp(nR_i) \rceil\}$ to $\{1, \dots, \lceil \exp(nR'_i) \rceil\}$, $i \in \mathcal{M}$. The random variables J_1, \dots, J_m , $J'_1, \dots, J'_m, X_\mathcal{M}^n$ are taken to be mutually independent. Define $\mathcal{J}_\mathcal{M}^0$ as the set of mappings $j_\mathcal{M} \in \mathcal{J}_\mathcal{M}$ for which there exists a $j'_\mathcal{M} \in \mathcal{J}'_\mathcal{M}$ such that $X_\mathcal{M}^n$ is ϵ_n -recoverable from $(X_i^n, j'_\mathcal{M}(j_\mathcal{M}(X_\mathcal{M}^n)))$ for every $i \in \mathcal{A}$. By the general source network coding theorem [3, Lemma 3.1.13 and Theorem 3.1.14], applied to the random mapping $J'_\mathcal{M}(J_\mathcal{M})$, it follows that for all sufficiently large n ,

$$\Pr(J_\mathcal{M} \in \mathcal{J}_\mathcal{M}^0) \geq 1 - \epsilon.$$

This, together with (26) and (29) in the proof of Theorem 5, imply that for a securely computable g there exist $j_\mathcal{M} \in \mathcal{J}_\mathcal{M}$ and $j'_\mathcal{M} \in \mathcal{J}'_\mathcal{M}$ for which the public communication $F'_\mathcal{M} \triangleq j'_\mathcal{M}(j_\mathcal{M})$ satisfies the aforementioned requirements. Finally, apply Lemma C2 with U, U', \mathcal{U}_0, V and d as in (35) but with $h = F'_\mathcal{M}$ and

$$r' = \left\lceil \exp \left[n \left(R_{CO}(\mathcal{A}) + \frac{\tau}{2} \right) \right] \right\rceil.$$

As in the proof above of part (i), a SK $K'_g = K_g'^{(n)}$ of rate

$$\frac{1}{n} \log r = H(X_\mathcal{M}|G) - R_{CO}(\mathcal{A}) - \tau = C(\mathcal{A}) - H(G) - \tau$$

which is nearly independent of $(F'_\mathcal{M}, G^n)$ (and, hence, satisfies (25)) exists for all n sufficiently large. \square

V. DISCUSSION

We obtain simple necessary and sufficient conditions for secure computability involving function entropy and ASK capacity. The latter is the largest rate of a SK for a new model in which side information is provided for use in only the recovery stage of SK generation. This model could be of independent interest. In particular, a function is securely computable if its entropy is less than ASK capacity of an associated secrecy model. The difference is shown to correspond to the maximum achievable rate of an ASK which is independent of the securely computed function and, together with it, forms an ASK of optimum rate. Also, a function that is securely computed by \mathcal{A} can be augmented to form a SK for \mathcal{A} of maximum rate.

Our results extend to functions defined on a block of symbols of *fixed* length in an obvious manner by considering larger alphabets composed of supersymbols of such length. However, they do not cover functions of symbols of increasing length (in n).

In our proof of Theorem 5, g was securely computed from omniscience at all the terminals in $\mathcal{A} \subseteq \mathcal{M}$ that was attained using noninteractive public communication. However, as Example 1 illustrates, omniscience is not necessary for the secure computation of g , and it is possible to make do with

communication of rate less than $R_{CO}(\mathcal{M})$ using an interactive protocol. A related unresolved question is: what is the minimum rate of public communication for secure computation?

A natural generalization of the conditions for secure computability of g by $\mathcal{A} \subseteq \mathcal{M}$ given here entails a characterization of conditions for the secure computability of multiple functions g_1, \dots, g_k by $\mathcal{A}_1, \dots, \mathcal{A}_k$ of \mathcal{M} , respectively. This unsolved problem, in general, will not permit omniscience for any $\mathcal{A}_i, i = 1, \dots, k$. For instance with $m = 2$, $\mathcal{A}_1 = \{1\}$, $\mathcal{A}_2 = \{2\}$, and X_1 and X_2 being independent, the functions $g_i(x_i) = x_i, i = 1, 2$, are securely computable trivially, but not through omniscience since, in this example, public communication is forbidden for the secure computation of g_1, g_2 .

APPENDIX A

The proof of Lemma 3 is based on [8, Lemma, Section 4], which is paraphrased first. Let the rvs Q and R take values in the finite set \mathcal{Q} and \mathcal{R} , respectively. For a stochastic matrix $W : \mathcal{Q} \rightarrow \mathcal{Q}$, let $\{\tilde{\mathcal{D}}_1, \dots, \tilde{\mathcal{D}}_l\}$ be the ergodic decomposition (into communicating classes) (cf. e.g., [12]) of \mathcal{Q} based on W . Let $\tilde{\mathcal{D}}^{(n)}$ denote a fixed ergodic class of \mathcal{Q}^n (the n -fold Cartesian product of \mathcal{Q}) on the basis of W^n (the n -fold product of W). Let $\mathcal{D}^{(n)}$ and $\mathcal{R}^{(n)}$ be any (nonempty) subsets of $\tilde{\mathcal{D}}^{(n)}$ and \mathcal{R}^n , respectively.

Lemma GK. [8] For $\tilde{\mathcal{D}}^{(n)}, \mathcal{D}^{(n)}, \mathcal{R}^{(n)}$ as above, assume that

$$\begin{aligned} \Pr(Q^n \in \mathcal{D}^{(n)} \mid R^n \in \mathcal{R}^{(n)}) &\geq \exp[-n\epsilon_n], \\ \Pr(R^n \in \mathcal{R}^{(n)} \mid Q^n \in \mathcal{D}^{(n)}) &\geq \exp[-n\epsilon_n], \end{aligned} \quad (\text{A1})$$

where $\lim_n \epsilon_n = 0$. Then (as stated in [8, bottom of p. 157]),

$$\frac{\Pr(Q^n \in \mathcal{D}^{(n)})}{\Pr(Q^n \in \tilde{\mathcal{D}}^{(n)})} \geq \exp[-n\kappa\epsilon_n \log^2 \epsilon_n], \quad (\text{A2})$$

for a (positive) constant κ that depends only on the pmf of (Q, R) and on W .

A simple consequence of (A2) is that for a given ergodic class $\tilde{\mathcal{D}}^{(n)}$ and disjoint subsets $\mathcal{D}_1^{(n)}, \dots, \mathcal{D}_t^{(n)}$ of it, and subsets $\mathcal{R}_1^{(n)}, \dots, \mathcal{R}_t^{(n)}$ (not necessarily distinct) of \mathcal{R}^n , such that $\mathcal{D}_{t'}^{(n)}, \mathcal{R}_{t'}^{(n)}, t' = 1, \dots, t$, satisfy (A1), then

$$t \leq \exp[n\kappa\epsilon_n \log^2 \epsilon_n]. \quad (\text{A3})$$

Note that the ergodic decomposition of Q^n on the basis of W^n for the specific choice

$$W(q|q') = \sum_{r \in \mathcal{R}} \Pr(Q = q \mid R = r) \Pr(R = r \mid Q = q'), \quad q, q' \in \mathcal{Q}$$

corresponds to the set of values of $\text{mcf}^n(Q, R)$ defined by (17) [8]. Next, pick $Q = Q_m, R = (Q_1, \dots, Q_{m-1})$, and define the stochastic matrix $W : \mathcal{Q} \rightarrow \mathcal{Q}$ by

$$\begin{aligned} W(q|q') &= \sum_{\alpha} \Pr(Q = q \mid \text{mcf}(Q_1, \dots, Q_{m-1}) = \alpha) \Pr(\text{mcf}(Q_1, \dots, Q_{m-1}) = \alpha \mid Q = q'), \\ q, q' &\in \mathcal{Q}. \end{aligned} \quad (\text{A4})$$

The ergodic decomposition of \mathcal{Q}^n on the basis of W^n (with W as in (A4)) will correspond to the set of values of $\text{mcf}^n(Q_1, \dots, Q_m)$, recalling (16). Since $\xi^{(n)}$ is ϵ -recoverable from $Q_i^n, i = 1, \dots, m$, note that

$$\xi'^{(n)} = \left(\xi^{(n)}, \text{mcf}^n(Q_1, \dots, Q_m) \right)$$

also is ϵ -recoverable in the same sense, recalling definition 7. This implies the existence of mappings $\xi_i'^{(n)}, i = 1, \dots, m$, satisfying

$$\Pr \left(\xi_1'^{(n)}(Q_1^n) = \dots = \xi_m'^{(n)}(Q_m^n) = \xi'^{(n)} \right) \geq 1 - \epsilon. \quad (\text{A5})$$

For each fixed value $c = (c_1, c_2)$ of $\xi'^{(n)}$, let

$$\begin{aligned} \mathcal{D}_c^{(n)} &= \left\{ q_m^n \in \mathcal{Q}_m^n : \xi_m'^{(n)}(q_m^n) = c \right\}, \\ \mathcal{R}_c^{(n)} &= \left\{ (q_1^n, \dots, q_{m-1}^n) \in \mathcal{Q}_1^n \times \dots \times \mathcal{Q}_{m-1}^n : \xi_i'^{(n)}(q_i^n) = c, i = 1, \dots, m-1 \right\}. \end{aligned}$$

Let $C(\epsilon)$ denote the set of c 's such that

$$\begin{aligned} \Pr \left(Q^n \in \mathcal{D}_c^{(n)} \mid R^n \in \mathcal{R}_c^{(n)} \right) &\geq 1 - \sqrt{\epsilon}, \\ \Pr \left(R^n \in \mathcal{R}_c^{(n)} \mid Q^n \in \mathcal{D}_c^{(n)} \right) &\geq 1 - \sqrt{\epsilon}. \end{aligned} \quad (\text{A6})$$

Then, as in [8, Proposition 1], it follows from (A5) that

$$\Pr \left(\xi'^{(n)} \in C(\epsilon) \right) \geq 1 - 4\sqrt{\epsilon}. \quad (\text{A7})$$

Next, we observe for each fixed c_2 , that the disjoint sets $\mathcal{D}_{c_1, c_2}^{(n)}$ lie in a fixed ergodic class of \mathcal{Q}^n (determined by c_2). Since (A6) are compatible with the assumption (A1) for all n sufficiently large, we have from (A3) that

$$\| \{c_1 : (c_1, c_2) \in C(\epsilon)\} \| \leq \exp[n\kappa\epsilon_n \log^2 \epsilon_n], \quad (\text{A8})$$

where κ depends on the pmf of (Q_1, \dots, Q_m) and W in (A4), and where $\lim_n \epsilon_n = 0$. Finally,

$$\begin{aligned} \frac{1}{n} H \left(\xi'^{(n)} \right) &= \frac{1}{n} H \left(\xi^{(n)}, \text{mcf}^n(Q_1, \dots, Q_m) \right) \\ &\leq H(\text{mcf}(Q_1, \dots, Q_m)) + \frac{1}{n} H \left(\xi^{(n)}, \mathbf{1} \left(\xi'^{(n)} \in C(\epsilon) \right) \mid \text{mcf}^n(Q_1, \dots, Q_m) \right) \\ &= H(\text{mcf}(Q_1, \dots, Q_m)) + \frac{1}{n} \\ &\quad + \frac{1}{n} H \left(\xi^{(n)} \mid \text{mcf}^n(Q_1, \dots, Q_m), \mathbf{1} \left(\xi'^{(n)} \in C(\epsilon) \right) \right) \\ &\leq H(\text{mcf}(Q_1, \dots, Q_m)) + \delta_n, \end{aligned}$$

where $\lim_n \delta_n = 0$ by (A7) and (A8). □

APPENDIX B

Considering first the achievability part, fix $\delta > 0$. From the result for a general source network [3, Theorem 3.1.14] it follows, as in the proof of [6, Proposition 1], that for $R_{\mathcal{M}} \in \mathcal{R}(\mathcal{A}', Z_{\mathcal{A}'})$ and all n sufficiently large, there exists a noninteractive communication $\mathbf{F}^{(n)} = (F_1^{(n)}, \dots, F_m^{(n)})$ with

$$\frac{1}{n} \log \|\mathbf{F}^{(n)}\| \leq \sum_{i=1}^m R_i + \delta,$$

such that $\mathcal{X}_{\mathcal{M}}^n$ is ϵ_n -recoverable from $(X_i^n, Z_i^n, \mathbf{F}^{(n)})$, $i \in \mathcal{A}'$. Therefore, $\{\text{mcf}((X_{\mathcal{M}t}, Z_{it})_{i \in \mathcal{A}'})\}_{t=1}^n$ is ϵ_n -recoverable from $(X_i^n, Z_i^n, \mathbf{F}^{(n)})$, $i \in \mathcal{A}'$. The last step takes recourse to Lemma C2 in Appendix C. Specifically, choose $U = U' = \{\text{mcf}((X_{\mathcal{M}t}, Z_{it})_{i \in \mathcal{A}'})\}_{t=1}^n$, $\mathcal{U}_0 = \mathcal{U}$, $V = \text{constant}$, $h = F^{(n)}$, $d = n[H(\text{mcf}((X_{\mathcal{M}}, Z_i)_{i \in \mathcal{A}'})) - \delta]$, whereby the hypothesis (C3) of Lemma C2 is satisfied for all n sufficiently large. Fixing

$$r' = \left\lceil \exp \left[n \left(\sum_{i=1}^m R_i + \delta \right) \right] \right\rceil,$$

Lemma C2 implies the existence of a ϕ , and thereby an ASK $K^{(n)} = \phi(\{\text{mcf}((X_{\mathcal{M}t}, Z_{it})_{i \in \mathcal{A}'})\}_{t=1}^n)$, of rate

$$\frac{1}{n} \log r = H(\text{mcf}((X_{\mathcal{M}}, Z_i)_{i \in \mathcal{A}'})) - \sum_{i=1}^m R_i - 3\delta.$$

In particular, we can choose

$$\sum_{i=1}^m R_i \leq R_{CO}(\mathcal{A}'; Z_{\mathcal{A}'}) + \frac{\delta}{2}.$$

Since δ was arbitrary, this establishes the achievability part.

We prove the converse part under either of the weaker conditions (7) or (8). Let $K = K^{(n)}(U_{\mathcal{M}}, X_{\mathcal{M}}^n, Z_{\mathcal{M}}^n)$ be an ϵ_n -ASK for \mathcal{A}' , achievable using observations of length n , randomization $U_{\mathcal{M}}$, public communication $\mathbf{F} = \mathbf{F}(U_{\mathcal{M}}, X_{\mathcal{M}}^n)$ and side information $Z_{\mathcal{M}}^n$. Then,

$$\frac{1}{n} H(K) \leq \frac{1}{n} H(K | \mathbf{F}) + \epsilon_n. \quad (\text{B1})$$

Let $K_u = K(u, X_{\mathcal{M}}^n, Z_{\mathcal{M}}^n)$ denote the random value of the ASK for a fixed $U_{\mathcal{M}} = u$. Since $(X_{\mathcal{M}}^n, K)$ is ϵ_n -recoverable from the rvs $(U_{\mathcal{M}}, X_{\mathcal{M}}^n, Z_{\mathcal{M}}^n)$ for each $i \in \mathcal{A}'$,

$$\begin{aligned} P_{U_{\mathcal{M}}}(\{u : (X_{\mathcal{M}}^n, K_u) \text{ is } \sqrt{\epsilon_n}\text{-recoverable from } (U_{\mathcal{M}} = u, X_{\mathcal{M}}^n, Z_{\mathcal{M}}^n) \text{ for each } i \in \mathcal{A}'\}) \\ \geq 1 - \sqrt{\epsilon_n}. \end{aligned} \quad (\text{B2})$$

Also, for each $U_{\mathcal{M}} = u$

$$\frac{1}{n} H(X_{\mathcal{M}}^n, K | U_{\mathcal{M}} = u) = \frac{1}{n} H(X_{\mathcal{M}}^n, K_u)$$

by independence of $U_{\mathcal{M}}$ and $(X_{\mathcal{M}}^n, Z_{\mathcal{M}}^n)$, and therefore, by Lemma 3, for u in the set in (B2),

$$\frac{1}{n}H(X_{\mathcal{M}}^n, K | U_{\mathcal{M}} = u) \leq H(\text{mcf}((X_{\mathcal{M}}, Z_i)_{i \in \mathcal{A}'})) + \delta_n, \quad (\text{B3})$$

for all n sufficiently large and where $\lim_n \delta_n = 0$. Then,

$$\frac{1}{n}H(U_{\mathcal{M}}, X_{\mathcal{M}}^n, K) \leq \frac{1}{n}H(U_{\mathcal{M}}) + H(\text{mcf}((X_{\mathcal{M}}, Z_i)_{i \in \mathcal{A}'})) + \delta_n + \sqrt{\epsilon_n} \log(|\mathcal{X}_{\mathcal{M}}||\mathcal{Z}_{\mathcal{M}}|), \quad (\text{B4})$$

by (B2) and (B3). The proof is now completed along the lines of [6, Lemma 2 and Theorem 3]. Specifically, denoting the set of positive integers $\{1, \dots, l\}$ by $[1, l]$,

$$\frac{1}{n}H(U_{\mathcal{M}}, X_{\mathcal{M}}^n, K) = \frac{1}{n}H(K | \mathbf{F}) + \sum_{i=1}^m R'_i + \frac{1}{n}H(U_{\mathcal{M}}),$$

where

$$R'_i = \frac{1}{n} \sum_{\nu: \nu \equiv i \pmod m} H(F_{\nu} | F_{[1, \nu-1]}) + \frac{1}{n}H(U_i, X_i^n | \mathbf{F}, K, U_{[1, i-1]}, X_{[1, i-1]}^n) - H(U_i). \quad (\text{B5})$$

Consider $B \not\subseteq \mathcal{M}$, $\mathcal{A}' \not\subseteq B$. For $j \in \mathcal{A}' \cap B^c$, we have

$$\begin{aligned} \frac{1}{n}H(U_B) + \frac{1}{n}H(X_B | X_{B^c}^n, Z_j^n) &= \frac{1}{n}H(U_B, X_B^n | U_{B^c}, X_{B^c}^n, Z_j^n) \\ &= \frac{1}{n}H(F_1, \dots, F_{rm}, K, U_B, X_B^n | U_{B^c}, X_{B^c}^n, Z_j^n). \end{aligned}$$

Furthermore, since K is ϵ_n -recoverable from $(\mathbf{F}, U_{B^c}, X_{B^c}^n, Z_j^n)$ and $H(F_{\nu} | U_{B^c}, X_{B^c}^n) = 0$ for $\nu \equiv i \pmod m$ with $i \in B^c$,

$$\begin{aligned} &\frac{1}{n}H(F_1, \dots, F_{rm}, K, U_B, X_B^n | U_{B^c}, X_{B^c}^n, Z_j^n) \\ &= \frac{1}{n} \sum_{\nu=1}^{rm} H(F_{\nu} | F_{[1, \nu-1]}, U_{B^c}, X_{B^c}^n, Z_j^n) + \frac{1}{n}H(K | U_{B^c}, X_{B^c}^n, Z_j^n, \mathbf{F}) \\ &\quad + \frac{1}{n} \sum_{i \in B} H(U_i, X_i^n | U_{B^c \cap [i+1, m]}, X_{B^c \cap [i+1, m]}^n, Z_j^n, \mathbf{F}, K, U_{[1, i-1]}, X_{[1, i-1]}^n) \\ &\leq \frac{1}{n} \sum_{i \in B} \left[\sum_{\nu: \nu \equiv i \pmod m} H(F_{\nu} | F_{[1, \nu-1]}) + H(U_i, X_i^n | \mathbf{F}, K, U_{[1, i-1]}, X_{[1, i-1]}^n) \right] + \frac{\epsilon_n \log |\mathcal{K}| + 1}{n} \\ &\leq \sum_{i \in B} R_i + H(U_B), \end{aligned} \quad (\text{B6})$$

where

$$R_i \triangleq \left(R'_i + \frac{\epsilon_n \log |\mathcal{K}| + 1}{n} \right), \quad i \in \mathcal{M}.$$

It follows from (B1) and (B4)-(B6) that

$$\frac{1}{n}H(K) \leq H(\text{mcf}((X_{\mathcal{M}}, Z_i)_{i \in \mathcal{A}'})) - \sum_{i=1}^m R_i + \left(\epsilon_n + \delta_n + \frac{\epsilon_n \log |\mathcal{K}| + 1}{n} + \sqrt{\epsilon_n} \log(|\mathcal{X}_{\mathcal{M}}||\mathcal{Z}_{\mathcal{M}}|) \right), \quad (\text{B7})$$

where $R_{\mathcal{M}} \in \mathcal{R}(\mathcal{A}', Z_{\mathcal{A}'})$ from (B6), and therefore

$$\sum_{i=1}^m R_i \geq R_{CO}(\mathcal{A}', Z_{\mathcal{A}'}). \quad (\text{B8})$$

Then, (B7), (B8) imply

$$\frac{1}{n}H(K) \leq C(\mathcal{A}', Z_{\mathcal{A}'}) + \left(\epsilon_n + \delta_n + \frac{\epsilon_n \log |\mathcal{K}| + 1}{n} + \sqrt{\epsilon_n} \log(|\mathcal{X}_{\mathcal{M}}||\mathcal{Z}_{\mathcal{M}}|) \right).$$

The proof is completed using the second part of (8) directly, or the second part of (7) in the manner of [6, Theorem 3]. This completes the converse part. \square

APPENDIX C

Our proof of achievability in Theorem 4 and sufficiency in Theorem 5 rely on a “balanced coloring lemma” in [1]; we state below a version of it from [6].

Lemma C1. [1, Lemma 3.1] *Let \mathcal{P} be any family of N pmfs on a finite set \mathcal{U} , and let $d > 0$ be such that $P \in \mathcal{P}$ satisfies*

$$P\left(\left\{u : P(u) > \frac{1}{d}\right\}\right) \leq \epsilon, \quad (\text{C1})$$

for some $0 < \epsilon < (1/9)$. Then the probability that a randomly selected mapping $\phi : \mathcal{U} \rightarrow \{1, \dots, r\}$ fails to satisfy

$$\sum_{i=1}^r \left| \sum_{u: \phi(u)=i} P(u) - \frac{1}{r} \right| < 3\epsilon,$$

simultaneously for each $P \in \mathcal{P}$, is less than $2Nr \exp\left(-\frac{\epsilon^2 d}{3r}\right)$.

In contrast to the application of Lemma C1 in [6, Lemma B.2], our mentioned proofs call for a balanced coloring of a set corresponding to a rv that differs from another rv for which probability bounds are used. However, both rvs agree with high probability when conditioned on a set of interest.

Consider rvs U, U', V with values in finite sets $\mathcal{U}, \mathcal{U}', \mathcal{V}$, respectively, where U' is a function of U , and a mapping $h : \mathcal{U} \rightarrow \{1, \dots, r'\}$. For $\lambda > 0$, let \mathcal{U}_0 be a subset of \mathcal{U} such that

- (i) $\Pr(U \in \mathcal{U}_0) > 1 - \lambda^2$;
- (ii) given $U \in \mathcal{U}_0, h(U) = j, U' = u', V = v$, there exists $u = u(u') \in \mathcal{U}_0$ satisfying

$$\Pr(U = u \mid h(U) = j, V = v, U \in \mathcal{U}_0) = \Pr(U' = u' \mid h(U) = j, V = v, U \in \mathcal{U}_0), \quad 1 \leq j \leq r', v \in \mathcal{V}. \quad (\text{C2})$$

Then the following holds.

Lemma C2. *Let the rvs U, U', V and the set \mathcal{U}_0 be as above. Further, assume that*

$$P_{UV} \left(\left\{ (u, v) : \Pr(U = u \mid V = v) > \frac{1}{d} \right\} \right) \leq \lambda^2. \quad (\text{C3})$$

Then, a randomly selected mapping $\phi : \mathcal{U}' \rightarrow \{1, \dots, r\}$ fails to satisfy

$$\sum_{j=1}^{r'} \sum_{v \in \mathcal{V}} \Pr(h(U) = j, V = v) \sum_{i=1}^r \left| \sum_{u' \in \mathcal{U}': \phi(u')=i} \Pr(U' = u' \mid h(U) = j, V = v) - \frac{1}{r} \right| < 14\lambda, \quad (\text{C4})$$

with probability less than $2rr'|\mathcal{V}| \exp\left(-\frac{c\lambda^3 d}{rr'}\right)$ for a constant $c > 0$.

Proof: Using the condition (i) in the definition of \mathcal{U}_0 , the left side of (C4) is bounded above by

$$2\lambda^2 + \sum_{j=1}^{r'} \sum_{v \in \mathcal{V}} \Pr(h(U) = j, V = v, U \in \mathcal{U}_0) \sum_{i=1}^r \left| \sum_{u' \in \mathcal{U}': \phi(u')=i} \Pr(U' = u' \mid h(U) = j, V = v, U \in \mathcal{U}_0) - \frac{1}{r} \right|.$$

Therefore, it is sufficient to prove that

$$\sum_{j=1}^{r'} \sum_{v \in \mathcal{V}} \Pr(h(U) = j, V = v, U \in \mathcal{U}_0) \sum_{i=1}^r \left| \sum_{u' \in \mathcal{U}': \phi(u')=i} \Pr(U' = u' \mid h(U) = j, V = v, U \in \mathcal{U}_0) - \frac{1}{r} \right| < 12\lambda, \quad (\text{C5})$$

with probability greater than $1 - 2rr'|\mathcal{V}| \exp\left(-\frac{c\lambda^3 d}{rr'}\right)$ for a constant $c > 0$.

Let $q = P_V \left(\left\{ v \in \mathcal{V} : \Pr(U \in \mathcal{U}_0 \mid V = v) < \frac{1-\lambda^2}{3} \right\} \right)$. Then, since

$$\begin{aligned} 1 - \lambda^2 &\leq \Pr(U \in \mathcal{U}_0) \leq \sum_{v \in \mathcal{V} : \Pr(U \in \mathcal{U}_0 \mid V=v) < \frac{1-\lambda^2}{3}} \Pr(U \in \mathcal{U}_0 \mid V = v) P_V(v) + (1 - q) \\ &< \frac{1 - \lambda^2}{3} q + (1 - q), \end{aligned}$$

we get from the extremities above that

$$q < \frac{3\lambda^2}{2}. \quad (\text{C6})$$

For $u \in \mathcal{U}_0$ and $v \in \mathcal{V}$ satisfying

$$\Pr(U \in \mathcal{U}_0 \mid V = v) \geq \frac{1 - \lambda^2}{3}, \quad \Pr(U = u \mid V = v, U \in \mathcal{U}_0) > \frac{3}{d(1 - \lambda^2)}, \quad (\text{C7})$$

we have that

$$\Pr(U = u|V = v) > \frac{1}{d}.$$

Therefore, by (C6) and (C3), it follows that

$$\sum_{(u,v): u \in \mathcal{U}_0, \Pr(U=u|V=v, U \in \mathcal{U}_0) > \frac{3}{d(1-\lambda^2)}} \Pr(U = u, V = v) \leq \lambda^2 + q < \frac{5\lambda^2}{2},$$

which is the same as

$$\begin{aligned} & \sum_{j=1}^{r'} \sum_{v \in \mathcal{V}} \Pr(h(U) = j, V = v, U \in \mathcal{U}_0) \\ & \sum_{u \in \mathcal{U}_0: \Pr(U=u|V=v, U \in \mathcal{U}_0) > \frac{3}{d(1-\lambda^2)}} \Pr(U = u|h(U) = j, V = v, U \in \mathcal{U}_0) < \frac{5\lambda^2}{2}. \end{aligned} \quad (\text{C8})$$

The bound in (C8) will now play the role of [6, inequality (50), p. 3059] and the remaining steps of our proof, which are parallel to those in [6, Lemma B.2], are provided here for completeness.

Setting

$$D = \left\{ (j, v) : \sum_{u \in \mathcal{U}: \Pr(U=u|V=v, U \in \mathcal{U}_0) > \frac{3}{d(1-\lambda^2)}} \Pr(U = u|h(U) = j, V = v, U \in \mathcal{U}_0) \leq \frac{5\lambda}{2} \right\}, \quad (\text{C9})$$

we get that

$$\sum_{(j,v) \in D^c} \Pr(h(U) = j, V = v, U \in \mathcal{U}_0) < \lambda. \quad (\text{C10})$$

Next, defining

$$E = \left\{ (j, v) : \Pr(h(U) = j, V = v, U \in \mathcal{U}_0) \geq \frac{\lambda}{r'} \Pr(V = v, U \in \mathcal{U}_0) \right\}, \quad (\text{C11})$$

it holds for $(j, v) \in E$,

$$\Pr(U = u|h(U) = j, V = v, U \in \mathcal{U}_0) \leq \frac{r'}{\lambda} \Pr(U = u|V = v, U \in \mathcal{U}_0). \quad (\text{C12})$$

Also,

$$\begin{aligned} \sum_{(j,v) \in E^c} \Pr(h(U) = j, V = v, U \in \mathcal{U}_0) & < \frac{\lambda}{r'} \sum_{j=1}^{r'} \sum_{v \in \mathcal{V}} \Pr(V = v, U \in \mathcal{U}_0) \\ & \leq \lambda. \end{aligned} \quad (\text{C13})$$

Further, for $(j, v) \in E$, if

$$\Pr(U = u|h(U) = j, V = v, U \in \mathcal{U}_0) > \frac{3r'}{\lambda d(1-\lambda^2)} \quad (\text{C14})$$

then from (C12), we have

$$\Pr(U = u | V = v, U \in \mathcal{U}_0) > \frac{3}{d(1 - \lambda^2)}. \quad (\text{C15})$$

Therefore, recalling the conditions that define \mathcal{U}_0 in (C2), we have for $(j, v) \in E \cap D$ that

$$\begin{aligned} & \sum_{\substack{u' \in \mathcal{U}': \\ \Pr(U' = u' | h(U) = j, V = v, U \in \mathcal{U}_0) > \frac{3r'}{\lambda d(1 - \lambda^2)}}} \Pr(U' = u' | h(U) = j, V = v, U \in \mathcal{U}_0) \\ &= \sum_{\substack{u' \in \mathcal{U}': \\ \Pr(U = u(u') | h(U) = j, V = v, U \in \mathcal{U}_0) > \frac{3r'}{\lambda d(1 - \lambda^2)}}} \Pr(U = u(u') | h(U) = j, V = v, U \in \mathcal{U}_0) \\ &= \sum_{\substack{u \in \mathcal{U}: \\ \Pr(U = u | h(U) = j, V = v, U \in \mathcal{U}_0) > \frac{3r'}{\lambda d(1 - \lambda^2)}}} \Pr(U = u | h(U) = j, V = v, U \in \mathcal{U}_0) \\ &\leq \frac{5\lambda}{2}, \end{aligned} \quad (\text{C16})$$

where second equality is by (C2), and the previous inequality is by (C14), (C15) and (C9). Also, using (C10), (C13), we get

$$\sum_{(j,v) \in E \cap D} \Pr(h(U) = j, V = v, U \in \mathcal{U}_0) \geq 1 - 2\lambda. \quad (\text{C17})$$

Now, the left side of (C5) is bounded, using (C17), as

$$\begin{aligned} & \sum_{j=1}^{r'} \sum_{v \in \mathcal{V}} \Pr(h(U) = j, V = v, U \in \mathcal{U}_0) \\ & \quad \sum_{i=1}^r \left| \sum_{u' \in \mathcal{U}': \phi(u') = i} \Pr(U' = u' | h(U) = j, V = v, U \in \mathcal{U}_0) - \frac{1}{r} \right| \\ & \leq 4\lambda + \sum_{(j,v) \in E \cap D} \Pr(h(U) = j, V = v, U \in \mathcal{U}_0) \\ & \quad \sum_{i=1}^r \left| \sum_{u' \in \mathcal{U}': \phi(u') = i} \Pr(U' = u' | h(U) = j, V = v, U \in \mathcal{U}_0) - \frac{1}{r} \right|. \end{aligned} \quad (\text{C18})$$

Using (C16), the family of pmfs $\{\Pr(U' = (\cdot) | h(U) = j, V = v, U \in \mathcal{U}_0), (j, v) \in E \cap D\}$ satisfies the hypothesis (C1) of Lemma C1 with d replaced by $\frac{\lambda(1-\lambda^2)d}{3r'}$ and ϵ replaced by $5\lambda/2$; assume that $0 < \lambda < 2/45$ so as to meet the condition following (C1). The mentioned family consists of at most

$r'|\mathcal{V}|$ pmfs. Therefore, using Lemma C1,

$$\sum_{j=1}^{r'} \sum_{v \in \mathcal{V}} \Pr(h(U) = j, V = v, U \in \mathcal{U}_0) \\ \sum_{i=1}^r \left| \sum_{u' \in \mathcal{U}': \phi(u')=i} \Pr(U' = u' \mid h(U) = j, V = v, U \in \mathcal{U}_0) - \frac{1}{r} \right| < \frac{23\lambda}{2}$$

with probability greater than

$$1 - 2rr'|\mathcal{V}| \exp\left(-\frac{25\lambda^3(1-\lambda^2)d}{36rr'}\right) \geq 1 - 2rr'|\mathcal{V}| \exp\left(-\frac{c\lambda^3d}{rr'}\right),$$

for a constant c . This completes the proof of (C5), and thereby the lemma. \square

ACKNOWLEDGEMENTS

The authors thank Sirin Nitinawarat for helpful discussions.

REFERENCES

- [1] R. Ahlswede and I. Csiszár, “Common randomness in information theory and cryptography—part i: Secret sharing,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 1121–1132, 1993.
- [2] R. Ahlswede and I. Csiszár, “On the oblivious transfer capacity,” *IEEE International Symposium on Information Theory (ISIT)*, pp. 2061–2064, 2007.
- [3] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless channels*. Academic Press, 1981.
- [4] I. Csiszár, “Almost independence and secrecy capacity,” *Prob. Pered. Inform.*, vol. 32, no. 1, pp. 48–57, 1996.
- [5] I. Csiszár and P. Narayan, “Common randomness and secret key generation with a helper,” *IEEE Trans. Inform. Theory*, vol. 46, pp. 344–366, March 2000.
- [6] I. Csiszár and P. Narayan, “Secrecy capacities for multiple terminals,” *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3047–3061, 2004.
- [7] —, “Secrecy capacities for multiterminal channel models,” *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2437–2452, 2008.
- [8] P. Gács and J. Körner, “Common information is far less than mutual information,” *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 149–162, 1973.
- [9] R. G. Gallager, “Finding parity in a simple broadcast network,” *IEEE Trans. Inform. Theory*, vol. 34, no. 2, pp. 176–180, 1988.
- [10] A. Giridhar and P. Kumar, “Computing and communicating functions over sensor networks,” *IEEE Journ. on Select. Areas in Commun.*, vol. 23, no. 4, pp. 755–764, 2005.
- [11] J. Körner and K. Marton, “How to encode the modulo-two sum of binary sources,” *IEEE Trans. Inform. Theory*, vol. 25, no. 2, pp. 219–221, 1979.
- [12] M. Loeve, *Probability Theory*. Van Nostrand New York, pp. 157 and 28–42, 1955.
- [13] N. Ma, P. Ishwar and P. Gupta, “Information-theoretic bounds for multiround function computation in collocated networks,” *IEEE International Symposium on Information Theory (ISIT)*, pp. 2306–2310, 2009.
- [14] U. M. Maurer, “Secret key agreement by public discussion from common information,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 733–742, May 1993.

- [15] U. M. Maurer, *Communications and Cryptography: Two sides of One Tapestry*, R.E. Blahut et al., Eds. ed. Norwell, MA: Kluwer, 1994, ch. 26, pp. 271–285.
- [16] A. Nascimento and A. Winter, “On the oblivious transfer capacity of noisy correlations,” *IEEE International Symposium on Information Theory (ISIT)*, pp. 1871–1875, 2009.
- [17] A. Orlitsky and A. El Gamal, “Communication with secrecy constraints,” *ACM Symp. on Theory of Computing (STOC)*, pp. 217–224, 1984.
- [18] A. Orlitsky and J. R. Roche, “Coding for computing,” *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 903–917, 2001.
- [19] A. D. Wyner, “Recent results in the shannon theory,” *IEEE Trans. Inform. Theory*, vol. 20, pp. 2–10, Jan 1974.
- [20] A. C. Yao, “Some complexity questions related to distributive computing,” *ACM Symp. on Theory of Computing (STOC)*, pp. 209–213, May 1979.
- [21] C. Ye, “Information theoretic generation of multiple secret keys,” *PhD thesis, Dept. Elect. and Compt. Eng., University of Maryland, College Park*, 2005.
- [22] C. Ye and P. Narayan, “Secret key and private key constructions for simple multiterminal source models,” *Proc. Int. Symp. Inform. Theory*, pp. 2133–2137, Sept 2005.